International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Analyzing and Predicting Cryptocurrency Flow with Advanced ML Algorithms

Sai Ravi Teja Gangavarapu[1], Ashish Chopra[2], Jeisank Gourava[3], Gaurav Somani[4], Samruddhi Bansod[5], Kritika Sharma[6], Neha Tummala[7], Bhavan Ram Kodavati[8]

[1, 2, 3, 4, 5, 6, 7, 8] St. Xavier's College, Kolkata, India

ABSTRACT

Currently, cryptocurrencies have emerged as a global phenomenon in the financial sector, becoming one of the most traded financial instruments worldwide. Despite their popularity, cryptocurrencies are among the most complex and obscure financial instruments, presenting a significant challenge due to their high volatility. This paper aims to apply machine learning techniques to the index and constituents of cryptocurrency to predict and forecast their prices. Specifically, the goal is to predict and forecast the closing price of the cryptocurrency index 30 and nine cryptocurrency constituents using various machine learning algorithms and models. By doing so, the paper seeks to simplify trading in these currencies. Several machine learning techniques and algorithms were employed and compared to determine the best results. We believe our work will help mitigate the challenges faced by cryptocurrency investors. The findings can significantly impact cryptocurrency portfolio management and help monitor price fluctuations in the cryptocurrency market. Additionally, our approach was compared with existing state-of-the-art methods from the literature, demonstrating that our best method delivers superior and competitive results. These prediction and forecasting methods can help people better understand market trends, making it easier to trade in the complex and challenging world of cryptocurrency.

Keywords: Cryptocurrency constituents, Cryptocurrency index, close price, machine learning.

Introduction

Cryptocurrencies are a type of virtual currency that rely on cryptography for security. These decentralized, open-source currencies operate on a peer-to-peer network. Cryptocurrencies typically utilize complex cryptographic algorithms that require a network of computers to perform computationally intensive mathematical operations . With cryptography integrated into their design, cryptocurrencies offer a new form of economy characterized by low-cost, online, and anonymous transactions. A comprehensive list of over 2,175 cryptocurrencies and their prices is available at [CoinMarketCap](https://coinmarketcap.com). These currencies operate independently of government control, are unregulated, and exhibit high volatility . Consequently, they can experience drastic overnight devaluations and aggressive price swings driven by public perception, making risk assessment challenging. The rising prices of cryptocurrencies have made mining a lucrative business .

Bitcoin, introduced by Satoshi Nakamoto on October 31, 2008, is one of the most valuable and decentralized cryptocurrencies, accounting for approximately 35% of the total market capitalization . Bitcoin's key innovation, blockchain, addresses the problem of double spending and reduces the control of centralized entities in transactions. Blockchain is a technology that maintains a record of all financial and economic transactions made in any cryptocurrency through a network of computers linked in a peer-to-peer network. Essentially, it is a robust technology capable of maintaining permanent records of commercial transactions, asset transfers, contracts, financial records, and intellectual property . The blockchain is a public ledger composed of blocks, and any node on the Bitcoin network can process and clear transactions by posting them .

Despite significant fluctuations in cryptocurrency prices since 2016, public enthusiasm for investing in these virtual currencies remains strong. These currencies are now used in official cash flows and goods exchanges. Consequently, researchers have developed various physical models and techniques to analyze cryptocurrency prices and market spontaneity to support real decision-making processes . These techniques include dynamic topic modeling, machine learning, data mining, and text mining. Additionally, agent-based artificial financial markets and genetic programming for identifying technical patterns have been proposed to study the cryptocurrency market . Given the correlation between cryptocurrencies , studies have also examined the cross-correlation of price changes using random matrix theory and minimum spanning trees . In recent years, machine learning algorithms have been employed to generate abnormal profits by exploiting market inefficiencies . Moreover, machine learning approaches have been used to measure the involvement of cryptocurrencies like Bitcoin in illicit activities accurately .

This paper aims to predict and forecast the closing price of the cryptocurrency index cci30 and its constituents to help mitigate market risks. We collected and analyzed historical data from [CoinMarketCap](https://coinmarketcap.com) and applied various machine learning techniques using RapidMiner software . We focused on ensemble learning methods, gradient boosted trees models, neural network models, and K-Nearest Neighbor (K-NN) models, evaluating their performance using standard metrics from the literature. Due to the high volatility of cryptocurrencies, a single algorithm may not be sufficient for accurate prediction or forecasting, which motivated us to explore and compare four different models. These models allow us to observe cryptocurrency behavior and determine the most effective algorithms for predicting and forecasting closing prices.

Related Works

Cryptocurrency represents a new digital asset in finance, characterized by extreme volatility compared to nearly all other financial instruments. Due to this high volatility and frequent price fluctuations, there is a scarcity of literature addressing price prediction for cryptocurrencies.

Xiaolei et al. [15] proposed three models—SVM, RF, and Light Gradient Boosting Machine—to forecast cryptocurrency market trends by combining daily data from 42 types of cryptocurrencies with key economic indicators. The Light Gradient Boosting Machine outperformed the other models, achieving accuracies of 0.776 for 2 months, 0.881 for 2 weeks, and 0.762 for 2 days when the test set was a true subset of the training set. When the test set was not a subset of the training set, accuracies were 0.607 for 2 weeks and 0.476 for 2 days, with higher performance noted on the first day of the period.

Lahmiri et al. [16] employed deep learning techniques—deep learning neural network (DLNN) and generalized regression neural networks (GRNN)—to forecast the prices of Bitcoin, Digital Cash, and Ripple. The RMSE values for Bitcoin were 2.75 × 10^3 (DLNN) and 8.80 × 10^3 (GRNN), for Digital Cash were 19.2923 (DLNN) and 50.2418 (GRNN), and for Ripple were 0.0499 (DLNN) and 0.3115 (GRNN), indicating high RMSE for Bitcoin and Digital Cash but lower values for Ripple.

Kim et al. [17] analyzed user comments in online cryptocurrency communities to predict price and transaction fluctuations. For Bitcoin, the accuracy of price and transaction predictions were 50.538% and 48.387%, respectively, over 13 days. For Ethereum, the accuracies were 49.425% and 51.149%, and for Ripple, the price fluctuation accuracy was 63.200%, with transaction fluctuations not considered.

Greaves et al. [18] used transaction graph data to predict Bitcoin prices, employing baseline, logistic regression, SVM, and neural network models. The respective accuracies were 53.4%, 54.3%, 53.7%, and 55.1%.

Barnwal et al. [19] proposed a stacking approach with neural networks for cryptocurrency investment, using Bitcoin data from Quandl. They combined extreme gradient boosting, K-NN, and Light Gradient Boosting Machine as discriminative classifiers optimized over a neural network layer. Accuracies for various models were reported for April-May and June-July 2018 periods, with the stacked generalization method showing competitive performance.

McNanny et al. [20] employed RNN, LSTM, and ARIMA models to predict Bitcoin prices in USD. RNN and LSTM outperformed ARIMA, with RNN and LSTM models achieving accuracies and RMSE values of 50.25% (52.78%) and 5.45% (6.87%), respectively.

Bakar et al. [21] used the ARIMA forecasting method to predict Bitcoin exchange rates in a high volatility environment, achieving absolute percentage errors of 1.4% for September 2017 and 9.3% for October 2017, with a mean absolute percentage error of 5.36%.

Rebane et al. [22] presented an ARIMA model and seq2seq recurrent deep multi-layer neural network (seq2seq) using various input types. Cumulative errors for seq2seq A, seq2seq B, seq2seq C, and ARIMA were 1.00, 0.89, 0.45, and 1.73, respectively. Visual comparisons over 40 days showed deviations from true values for all models.

Materials and Methods

2.1 Dataset

Our dataset, sourced from [CoinMarketCap](https://coinmarketcap.com), includes daily data for seven-day weeks. The constituents and predictive models utilized are detailed in Table 1. We chose seven attributes (see Table 2) and divided our data into training and testing subsets.

Using RapidMiner, we developed various models to predict the closing prices of the constituents and CCI30 for January 2019 based on historical data. For the gradient-boosted tree and neural network models, we split the data into training and testing datasets (refer to Table 2). The dataset for the ensemble learning method is shown in Table 4. The K-NN model exclusively involves the training phase for all cryptocurrencies, with no testing phase, as our goal is to forecast values for January 2019 (see Table 3).

2.2 Performance Metrics

The performance metrics used in this paper include root mean squared error (RMSE), prediction trend accuracy, absolute error, relative error, squared error, correlation, and squared correlation. These metrics were obtained using the "Performance (Regression)" and "Forecasting Performance" operators.

- Root Mean Squared Error (RMSE): Measures the differences between the values predicted by the model and the actual values.
- Prediction Trend Accuracy: Measures the average number of times a regression prediction correctly predicted the trend of the regression.
- Absolute Error: The average absolute deviation of the prediction from the actual value, where the label attribute values are the actual values.
- Relative Error: The average of the absolute deviation of the prediction from the actual value divided by the actual value.
- Squared Error: Selects the model with the smallest average squared error value.
- Correlation: Returns the correlation coefficient between the "label" and "prediction" attributes.
- Squared Correlation: Returns the squared correlation coefficient between the "label" and "prediction" attributes.

2.1 Methodology

Machine learning approaches play a critical role in the finance domain, particularly for predicting the prices of financial instruments, including cryptocurrencies. From managing cryptocurrency portfolios to predicting price fluctuations in cryptocurrency transactions, machine learning has proven to be one of the most effective techniques. Integrating these techniques into business intelligence systems enables real-life decision-making. However, predicting and analyzing cryptocurrency prices is challenging due to their high volatility. As discussed in the literature review, various ML-based approaches have been proposed to address these challenges.

In this paper, we used RapidMiner, a widely used software that supports all steps of the data mining process. RapidMiner allows for data analysis through graphs, plots, charts, and tables, making it easy to visualize outputs and compare attributes and models. For predicting and forecasting the future closing prices of cryptocurrencies, it is essential to train the machine on the given dataset. Models are created using different algorithms, which are then used to accomplish the prediction/forecasting tasks.

2.1.1 Predictive Model: Gradient Boosted Trees

The gradient boosted trees model is particularly advantageous for price prediction for several reasons. Firstly, data normalization is not required, as the model is sensitive to the arithmetic range of data and features. Secondly, it is a scalable machine learning model due to its construction process. Lastly, it is a rule-based learning method. Many studies have successfully employed gradient boosted trees for predicting sales and cryptocurrency prices.

Gradient boosted trees combine either regression or classification tree models, using a forward learning ensemble method to gradually improve predictive results. To predict the closing price of cryptocurrencies for January 2019, we used the attributes listed in Table 2 and historical price data from Table 3. RapidMiner was used to optimize the model parameters through various permutations. The parameter "number of trees" was tuned to 500, while other parameters were set to default values. The "Performance (Regression)" and "Forecasting Performance" operators determined the performance of the testing dataset. Chart 1(a) illustrates the performance of the gradient boosted trees model using all metrics mentioned in Section 3.2 for all nine constituents; the actual values are provided in supplement S1. Figure 1 shows the model, and Figure 2 (a)-(i) presents graphs comparing the original and predicted closing prices for January 2019. A portion of the gradient boosted tree model leaf is shown in supplement S2.

2.1 Methodology

Machine learning approaches play a critical role in the finance domain, particularly for predicting the prices of financial instruments, including cryptocurrencies. From managing cryptocurrency portfolios to predicting price fluctuations in cryptocurrency transactions, machine learning has proven to be one of the most effective techniques. Integrating these techniques into business intelligence systems enables real-life decision-making. However, predicting and analyzing cryptocurrency prices is challenging due to their high volatility. As discussed in the literature review, various ML-based approaches have been proposed to address these challenges.

In this paper, we used RapidMiner, a widely used software that supports all steps of the data mining process. RapidMiner allows for data analysis through graphs, plots, charts, and tables, making it easy to visualize outputs and compare attributes and models. For predicting and forecasting the future closing prices of cryptocurrencies, it is essential to train the machine on the given dataset. Models are created using different algorithms, which are then used to accomplish the prediction/forecasting tasks.

2.1.1 Predictive Model: Gradient Boosted Trees

The gradient boosted trees model is particularly advantageous for price prediction for several reasons. Firstly, data normalization is not required, as the model is sensitive to the arithmetic range of data and features. Secondly, it is a scalable machine learning model due to its construction process. Lastly, it is a rule-based learning method. Many studies have successfully employed gradient boosted trees for predicting sales and cryptocurrency prices.

Gradient boosted trees combine either regression or classification tree models, using a forward learning ensemble method to gradually improve predictive results. To predict the closing price of cryptocurrencies for January 2019, we used the attributes listed in Table 2 and historical price data from Table 3. RapidMiner was used to optimize the model parameters through various permutations. The parameter "number of trees" was tuned to 500, while other parameters were set to default values. The “Performance (Regression)” and “Forecasting Performance” operators determined the performance of the testing dataset. Chart 1(a) illustrates the performance of the gradient boosted trees model using all metrics mentioned in Section 3.2 for all nine constituents; the actual values are provided in supplement S1. Figure 1 shows the model, and Figure 2 (a)-(i) presents graphs comparing the original and predicted closing prices for January 2019. A portion of the gradient boosted tree model leaf is shown in supplement S2.

2.1.2 Predictive Model: Neural Net

An artificial neural network (ANN) is a mathematical and computational model inspired by the structure and function of biological neural networks. ANNs consist of interconnected artificial neurons and process information using a connectionist approach to computation. Neural networks can model complex relationships between inputs and outputs or identify patterns in data to predict cryptocurrency prices.

We used the same model described in [25], illustrated in Figure 3. Historical prices of all seven attributes were input into the training dataset, while attributes for January 2019 were provided in the testing dataset. The "Set Role" operator set the attribute name as “Date” and the target role as “id”. A windowing operator created examples from the input data by windowing the value series. The "series representation" parameter was set to "encode_series_by_examples," the "window size" and "step size" parameters were both set to 1. "Create single attribute" and "create label" options were checked to predict future closing prices based on past close prices. The testing side did not create a label attribute, as it would be predicted by the model. The "Horizon" parameter, representing the distance between the last window value and the value to predict, remained at its default state (1).

The sliding window validation operator enclosed sliding windows of training and testing to estimate the performance of the prediction operator. The "training window width" was set to 4, the "training window step size" to 1, and the "test window width" to 4. The sliding window validation operator, a subprocess with training and testing phases, was provided with a neural net model during training. The training parameters included 500 training cycles, a learning rate of 0.03, and a momentum of 0.9 to avoid local maxima and smooth optimization directions. The testing side included an “Apply Model” operator with a “Forecasting Performance” or “Performance (Regression)” operator. Chart 1(b) illustrates the performance of the neural net model using all metrics mentioned in Section 3.2 for all nine constituents; actual values are provided in supplement S3. Figure 4 (a)-(i) shows comparative graphs of original and predicted closing prices for the constituents.

An artificial neural network (ANN) is a mathematical and computational model inspired by the structure and function of biological neural networks. ANNs consist of interconnected artificial neurons and process information using a connectionist approach to computation. Neural networks can model complex relationships between inputs and outputs or identify patterns in data to predict cryptocurrency prices.

We used the same model described in [25], illustrated in Figure 3. Historical prices of all seven attributes were input into the training dataset, while attributes for January 2019 were provided in the testing dataset. The "Set Role" operator set the attribute name as “Date” and the target role as “id”. A windowing operator created examples from the input data by windowing the value series. The "series representation" parameter was set to "encode_series_by_examples," the "window size" and "step size" parameters were both set to 1. "Create single attribute" and "create label" options were checked to predict future closing prices based on past close prices. The testing side did not create a label attribute, as it would be predicted by the model. The "Horizon" parameter, representing the distance between the last window value and the value to predict, remained at

its default state (1).

The sliding window validation operator enclosed sliding windows of training and testing to estimate the performance of the prediction operator. The "training window width" was set to 4, the "training window step size" to 1, and the "test window width" to 4. The sliding window validation operator, a subprocess with training and testing phases, was provided with a neural net model during training. The training parameters included 500 training cycles, a learning rate of 0.03, and a momentum of 0.9 to avoid local maxima and smooth optimization directions. The testing side included an “Apply Model” operator with a “Forecasting Performance” or “Performance (Regression)” operator. Chart 1(b) illustrates the performance of the neural net model using all metrics mentioned in Section 3.2 for all nine constituents; actual values are provided in supplement S3. Figure 4 (a)-(i) shows comparative graphs of original and predicted closing prices for the constituents.

Predictive Model: Ensemble Learning Method

In an ensemble learning method multiple machine learning algorithms or learners are strategically generated and combined together in order to solve one particular computational intelligence problem. Using ensemble learning method, we can construct set of models and combine them through voting process, as a result of which, the likelihood of an unfortunate selection of a bad model for prediction can be reduced. This method helps to overcome the biases and error rates of the individual (weak) models by combining them through creating a strong learner by uniting some weak learners. The training dataset plays the role of most effective contributor to the error in the model so we have taken a large training dataset with seven attributes and close price as the label attribute. In recent years, several ensemble learning techniques have been proposed to find and predict the prices of cryptocurrencies [12,30].

2.1 Methodology

Machine learning approaches play a critical role in finance, particularly in predicting the prices of financial instruments, including cryptocurrencies. These techniques are integral to business intelligence systems for real-life decision-making. However, due to the high volatility of cryptocurrencies, predicting their prices is particularly challenging. Various ML-based approaches have been proposed to address these challenges.

In this paper, we used RapidMiner, a comprehensive software tool that supports all steps of the data mining process. RapidMiner allows for data analysis through graphs, plots, charts, and tables, making it easy to visualize outputs and compare attributes and models. For predicting and forecasting the future closing prices of cryptocurrencies, it is essential to train the machine on the given dataset. Models are created using different algorithms to accomplish the prediction and forecasting tasks.

2.1.1 Predictive Model: Gradient Boosted Trees

The gradient boosted trees model is particularly advantageous for price prediction for several reasons. Firstly, it does not require data normalization as it is sensitive to the arithmetic range of data and features. Secondly, it is a scalable machine learning model. Lastly, it is a rule-based learning method. Many studies have successfully employed gradient boosted trees for predicting sales and cryptocurrency prices.

Gradient boosted trees combine regression or classification tree models, using a forward learning ensemble method to gradually improve predictive results. To predict the closing price of cryptocurrencies for January 2019, we used the attributes listed in Table 2 and historical price data from Table 3. RapidMiner was used to optimize the model parameters through various permutations. The parameter "number of trees" was tuned to 500, while other parameters were set to default values. The “Performance (Regression)” and “Forecasting Performance” operators determined the performance of the testing dataset. Chart 1(a) illustrates the performance of the gradient boosted trees model using all metrics mentioned in Section 3.2 for all nine constituents; the actual values are provided in supplement S1. Figure 1 shows the model, and Figure 2 (a)-(i) presents graphs comparing the original and predicted closing prices for January 2019. A portion of the gradient boosted tree model leaf is shown in supplement S2.

2.1.2 Predictive Model: Neural Net

An artificial neural network (ANN) is a mathematical and computational model inspired by the structure and function of biological neural networks. ANNs consist of interconnected artificial neurons and process information using a connectionist approach to computation. Neural networks can model complex relationships between inputs and outputs or identify patterns in data to predict cryptocurrency prices.

We used the same model described in [25], illustrated in Figure 3. Historical prices of all seven attributes were input into the training dataset, while attributes for January 2019 were provided in the testing dataset. The "Set Role" operator set the attribute name as “Date” and the target role as “id”. A windowing operator created examples from the input data by windowing the value series. The "series representation" parameter was set to "encode_series_by_examples," the "window size" and "step size" parameters were both set to 1. "Create single attribute" and "create label" options were checked to predict future closing prices based on past close prices. The testing side did not create a label attribute, as it would be predicted by the model. The "Horizon" parameter, representing the distance between the last window value and the value to predict, remained at its default state (1).

The sliding window validation operator enclosed sliding windows of training and testing to estimate the performance of the prediction operator. The "training window width" was set to 4, the "training window step size" to 1, and the "test window width" to 4. The sliding window validation operator, a subprocess with training and testing phases, was provided with a neural net model during training. The training parameters included 500 training cycles, a learning rate of 0.03, and a momentum of 0.9 to avoid local maxima and smooth optimization directions. The testing side included an “Apply Model” operator with a “Forecasting Performance” or “Performance (Regression)” operator. Chart 1(b) illustrates the performance of the neural net model using all metrics mentioned in Section 3.2 for all nine constituents; actual values are provided in supplement S3. Figure 4 (a)-(i) shows comparative graphs of original and predicted closing prices for the constituents.

2.1.3 Predictive Model: Ensemble Learning Method

In this paper, we created an ensemble learning method to achieve better results by employing multiple models together, as shown in Figure 5. The “Split Data” operator partitions our data into subsets, with parameters “partitions” and “sampling type.” The “partitions” parameter splits data for training and testing in a 0.6 to 0.4 ratio, and the “sampling type” is set to “linear sampling” to divide the dataset without changing the order of examples. The “Vote” operator, a subprocess, uses majority vote for classification or average for regression. We included gradient boosted trees, neural net, and relative regression operators inside the “Vote” operator.

For the gradient boosted trees model, the number of trees parameter was set to 500, with other parameters at default. The neural net model had 500 training cycles, a learning rate of 0.3, and momentum set to 0.2. The relative regression operator learned a regression model for predictions relative to another attribute value, using a “Linear Regression” model with default parameters. Multiple relative regression learners minimized error through a voting process. Finally, the “Apply Model” operator applied the models through the voting process with a “Performance (Regression)” or “Forecasting Performance” operator to determine performance metrics. Chart 1(c) illustrates the performance of the ensemble learning method using all metrics in Section 3.2 for all nine constituents; actual values are in supplement S4. Comparative graphs between original and predicted closing prices are shown in Figure 6 (a)-(i).

2.1.4 Predictive Model: K-NN

The K-NN (k-Nearest Neighbor) algorithm is a non-parametric method that compares an unknown dataset with the K training examples closest to the unknown example. This algorithm can be used for classification or regression, making it valuable for forecasting cryptocurrency prices. It has also been used for detecting covert cryptocurrency mining operations and managing location data of IoT service providers and users based on blockchain with smart contracts.

In this paper, we used a forecasting model similar to [33]. For forecasting the close prices of these currencies using the K-NN model, we considered the attributes “Date” and “Close Price” only, as shown in Figure 7. We used this model to forecast the close prices for January 2019. Residual analysis was performed to evaluate the time series model, indicating the difference between our forecasting method and actual values. Good forecasting methods yield uncorrelated residuals with a mean near zero. The average residuals for all cryptocurrencies and the index are shown in Table 5. Chart 1(d) illustrates the performance of the K-NN model using all metrics in Section 3.2 for all nine constituents; actual values are in supplement S5. Figure 8 (a)-(i) shows differences between original and forecasted close prices for January 2019. Figure 10 compares the predicted and forecasted values by all models.

2.2 Index

We used the “Cryptocurrencies Index 30 (cci30)” for prediction and forecasting from [CCI30](https://cci30.com). This rules-based index measures the overall growth and movements of the blockchain sector by tracking the thirty largest cryptocurrencies by market capitalization. Launched on January 1, 2017, its starting value was set at 100 on January 1, 2015.

We downloaded the OHLCV daily values of the index from [CCI30](https://cci30.com). To predict the daily close price of the index for January 2019, we used gradient boosted trees, ensemble learning methods, neural network models, and the K-NN algorithm in RapidMiner. The training and testing datasets for prediction are given in Table 6, while only the training phase was used for the K-NN model to forecast January 2019 close prices. The attributes considered, except volume and market capitalization, are listed in Table 5. The performance vectors of the models are shown in supplement S6, and Figure 9 (a)-(d) illustrates the comparison between original and predicted close prices for January 2019 from all models. Figure 10 (j) presents the comparison, including all models together.

Comparison with Previous Results

This section presents a detailed comparison among the results obtained by the four models in this paper and other state of the art methods.
A Gradient Boosting Decision Tree (GBDT) algorithm, Light Gradient Boosting Machine (LightGBM) is adopted in [15] to forecast the price trend of cryptocurrency market, where they have argued that, the robustness of LightGBM model works better than SVM and RF model. Using LightGBM model, maximum forecasting performance in the first category of training sets has accuracy of 0.905 for two weeks and in second category of training sets, the accuracy is 0.952 for two weeks. In Table 7, we have shown a comparison between the results obtained in this paper and that in [15], which shows that our accuracy is 0.924 using ensemble learning method, whereas in [15], the highest accuracy is 0.952 using LightGBM model. However, they have forecasted the price trend (falling, not falling) by combining daily data of 42 kinds of cryptocurrencies with key economy indicators, but we have predicted and forecasted the price of 9 cryptocurrencies and cci30. In paper [15], only six months daily trading data from January 1, 2018 to June 30, 2018 were collected and considered for forecasting, but in our paper, we have taken yearly data, as a result of which our result is slightly less than in [15].

Deep learning techniques have been employed to forecast the price of Bitcoin, Digital Cash and Ripple in [16], where they have demonstrated that long-short term memory neural network (LSTM) performs better than generalized regression neural networks. The RMSE obtained using deep learning LSTM networks for Bitcoin, Digital Cash and Ripple are 2.75×10^3 19.2923 and 0.0499. In Table 8, we have shown a comparison between the results obtained in this paper and that in [16], which depicts that the models and learners we have used for prediction and forecast, yielded much better results than the work done in [16]. The RMSE we have obtained by Gradient Boosted Trees for Bitcoin, Doge coin and NEM are 32.863, 0.000 and 0.001.
Non-linear deep learning methods (LSTM and RNN) and ARIMA model are employed in [20] to predict the direction of Bitcoin prices in USD with good accuracy. It has been demonstrated that, LSTM and RNN have performed better than ARIMA model. LSTM gives accuracy and RMSE of 52.78% and 6.87%, while RNN gives accuracy and RMSE of 50.25% and 5.45%; for ARIMA model, the values are 50.05% and 53.74% respectively. In Table 9, we have shown a comparison between the results obtained in this paper and that in [20]. It is seen by comparing that, our maximum accuracy and RMSE are much better than it is in paper [20]. Our highest accuracy and RMSE using gradient boosted trees model are 0.900 and 0.001, and 0.924 and 0.002 using ensemble learning method.

4. Conclusion

In this article, we have presented four different models—Gradient Boosted Trees, Neural Network, Ensemble Learning, and K-Nearest Neighbor (K-NN)—to predict and forecast the closing prices of nine cryptocurrency constituents and the Cryptocurrency Index 30 (cci30) using machine learning techniques. Our models demonstrate strong performance in predicting the closing prices of cryptocurrencies, which can be extremely beneficial for public, private, and government organizations. These models help in understanding the trends and patterns of these currencies, providing valuable insights for decision-making.

However, our analysis revealed that the K-NN model was less effective in forecasting compared to the other models. This reduced effectiveness is likely due to the presence of noisy random features and the extreme volatility inherent in cryptocurrency markets. Despite this, our comparative analysis with state-of-the-art models from existing literature indicates that our models deliver better and more competitive performance.

We believe that our models will aid individuals and organizations in observing and understanding cryptocurrency market trends, thus helping them make informed decisions when selecting and trading their desired cryptocurrencies. Our work has the potential to contribute significantly to the field of cryptocurrency prediction and forecasting, offering a robust tool for navigating the highly volatile and dynamic cryptocurrency market.

Future Work

To further enhance the robustness and accuracy of our models, future work can focus on several key areas:

1. Feature Engineering: Exploring advanced feature engineering techniques to reduce noise and enhance the signal in the data, improving model performance.

2. Model Optimization: Continuously optimizing model parameters and experimenting with newer algorithms to achieve even better prediction accuracy.

3. Real-Time Data Integration: Incorporating real-time data streams to improve the responsiveness and applicability of the models in live trading scenarios.

4. Multivariate Analysis: Expanding the analysis to include multivariate time series models that consider multiple influencing factors simultaneously.

5. Risk Assessment: Developing integrated risk assessment frameworks that complement the predictive models, providing a comprehensive tool for cryptocurrency investors.

6. Cross-Market Analysis: Extending the models to analyze and predict relationships between different cryptocurrency markets and other financial markets to understand broader economic impacts.

7. User-Friendly Applications: Creating user-friendly applications and dashboards that leverage these predictive models, making them accessible to a broader audience including non-experts.

By addressing these areas, we can continue to refine our approach and offer even more powerful tools for cryptocurrency market analysis and decision-making.

Implications

The successful application of these models has several important implications:

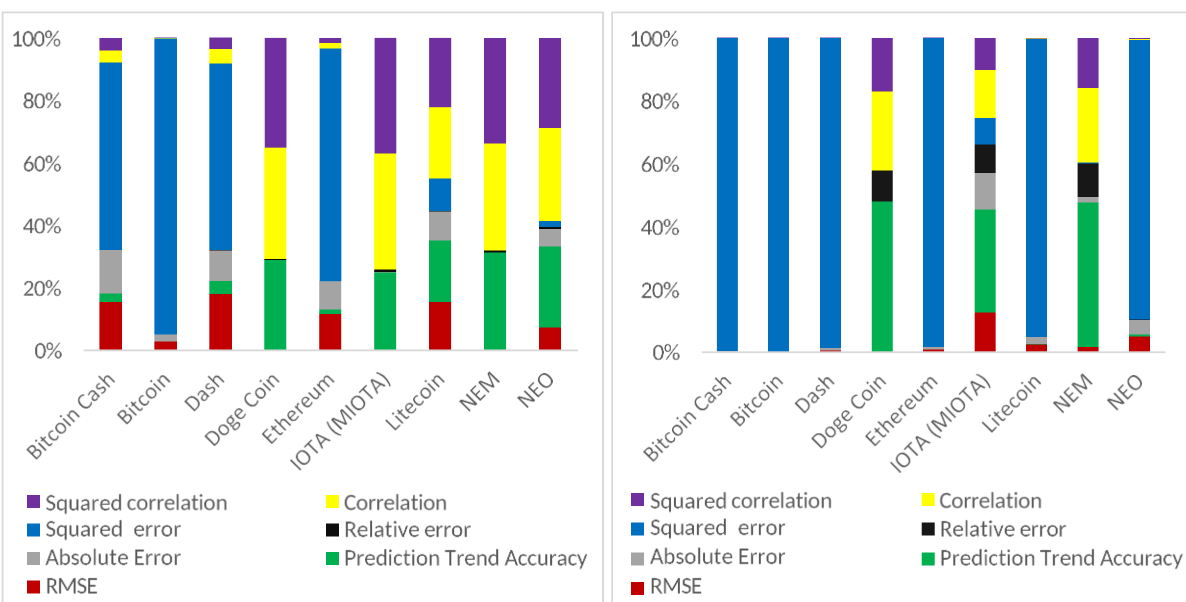
1. Investment Strategies: Investors can use the models to develop more informed and strategic investment plans, reducing risk and maximizing returns.

2. Market Regulation: Government and regulatory bodies can utilize these models to monitor and understand market dynamics, potentially informing regulatory decisions.

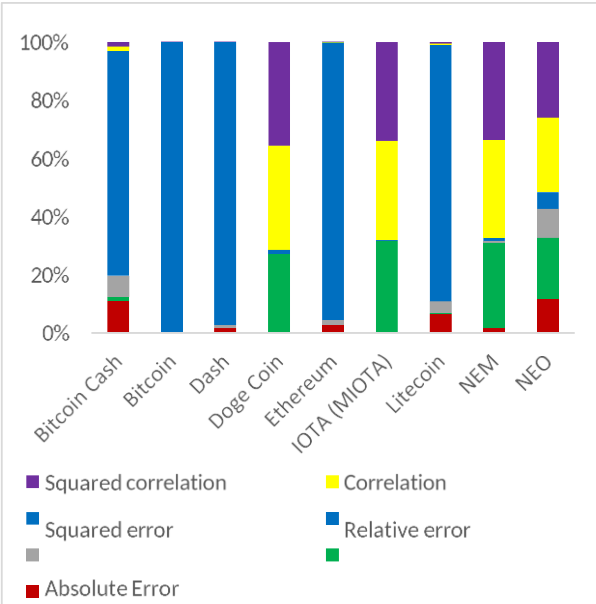
3. Financial Education: Educators and researchers can use the models as teaching tools to explain complex financial concepts and the impact of machine learning in finance.

Chart 1: Performance measurements

(a) (b)



(c)



(d)

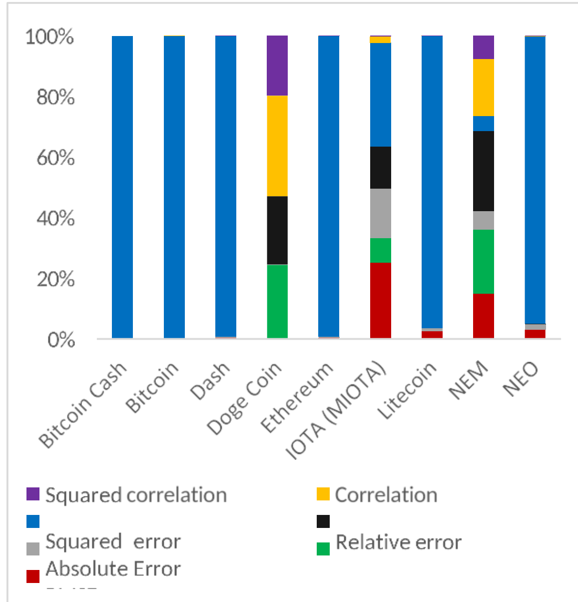


Chart 1: Performance measures of (a) Gradient Boosted Trees model (b) Neural Network model (c) Ensemble learning method (d) K-NN model.

Figures and Captions List

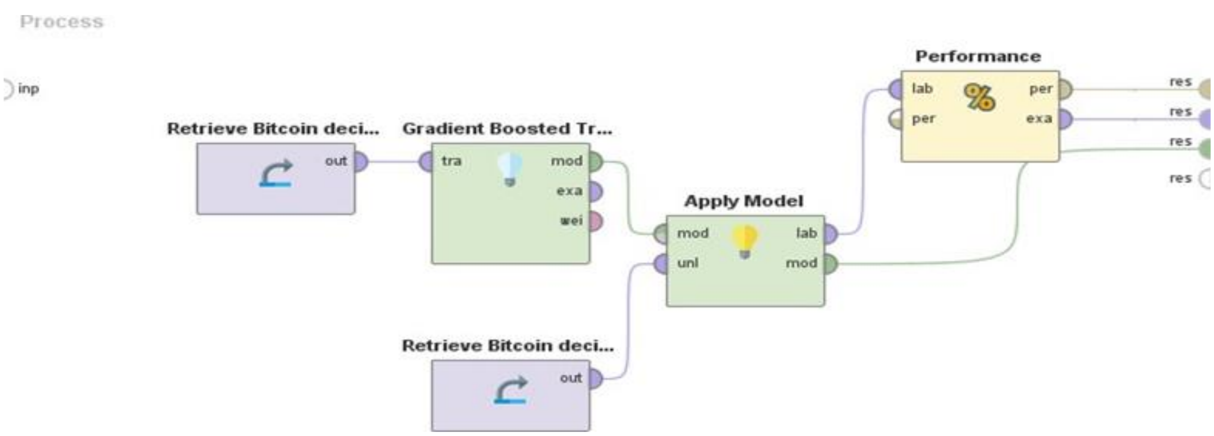


Figure 1. Gradient boosted trees model for predicting constituents and index of cryptocurrencies.

(a)

(b)

(c)

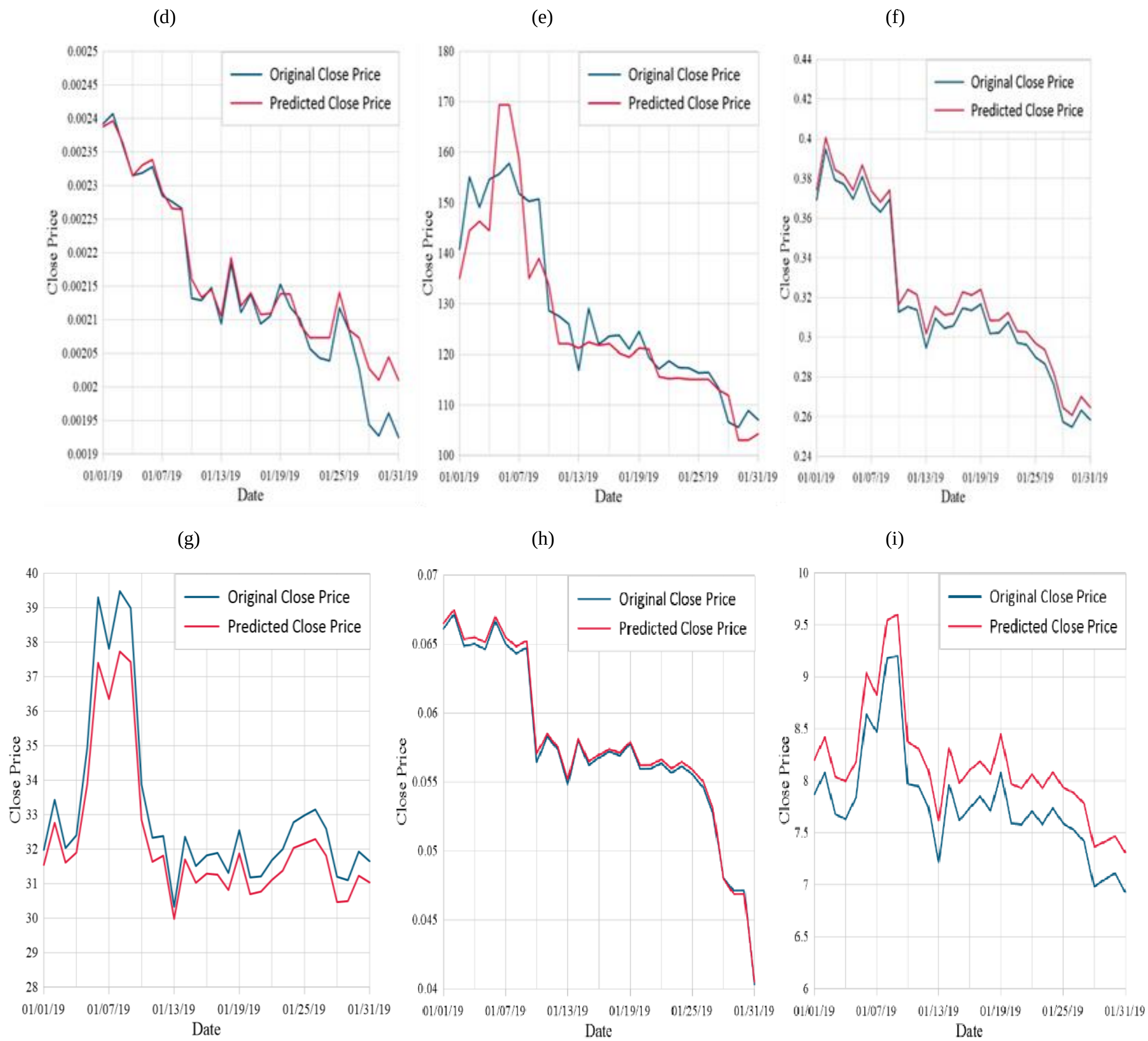


Figure 2. Comparison between original and predicted close price obtained from RapidMiner using model gradient boosted trees for the month January 2019 (a) of constituent Bitcoin Cash. (b) of constituent Bitcoin. (c) of constituent Dash. (d) of constituent Dogecoin (DOGE). (e) of constituent Ethereum. (f) of constituent IOTA (MIOTA). (g) of constituent Litecoin. (h) of constituent NEM. (i) of constituent NEO.

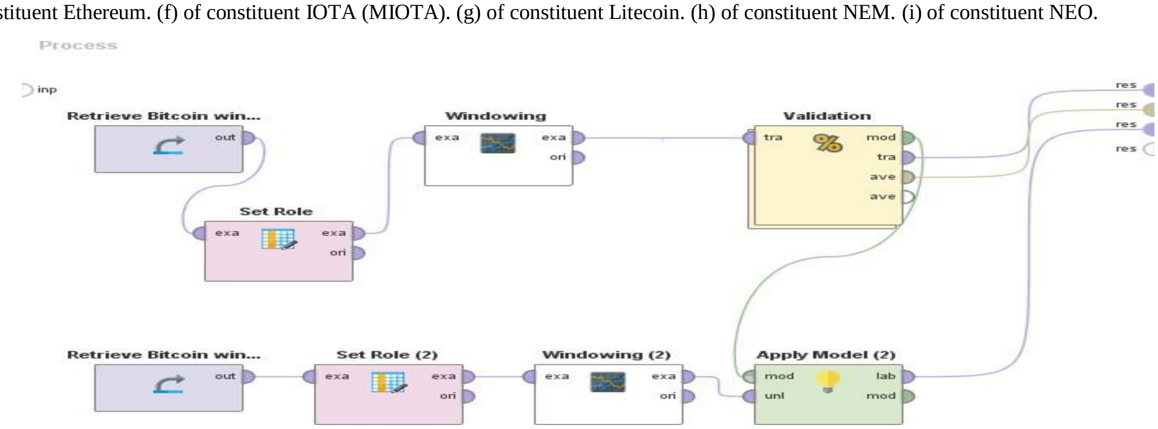


Figure 3. Neural net model for predicting constituents and index of cryptocurrencies.

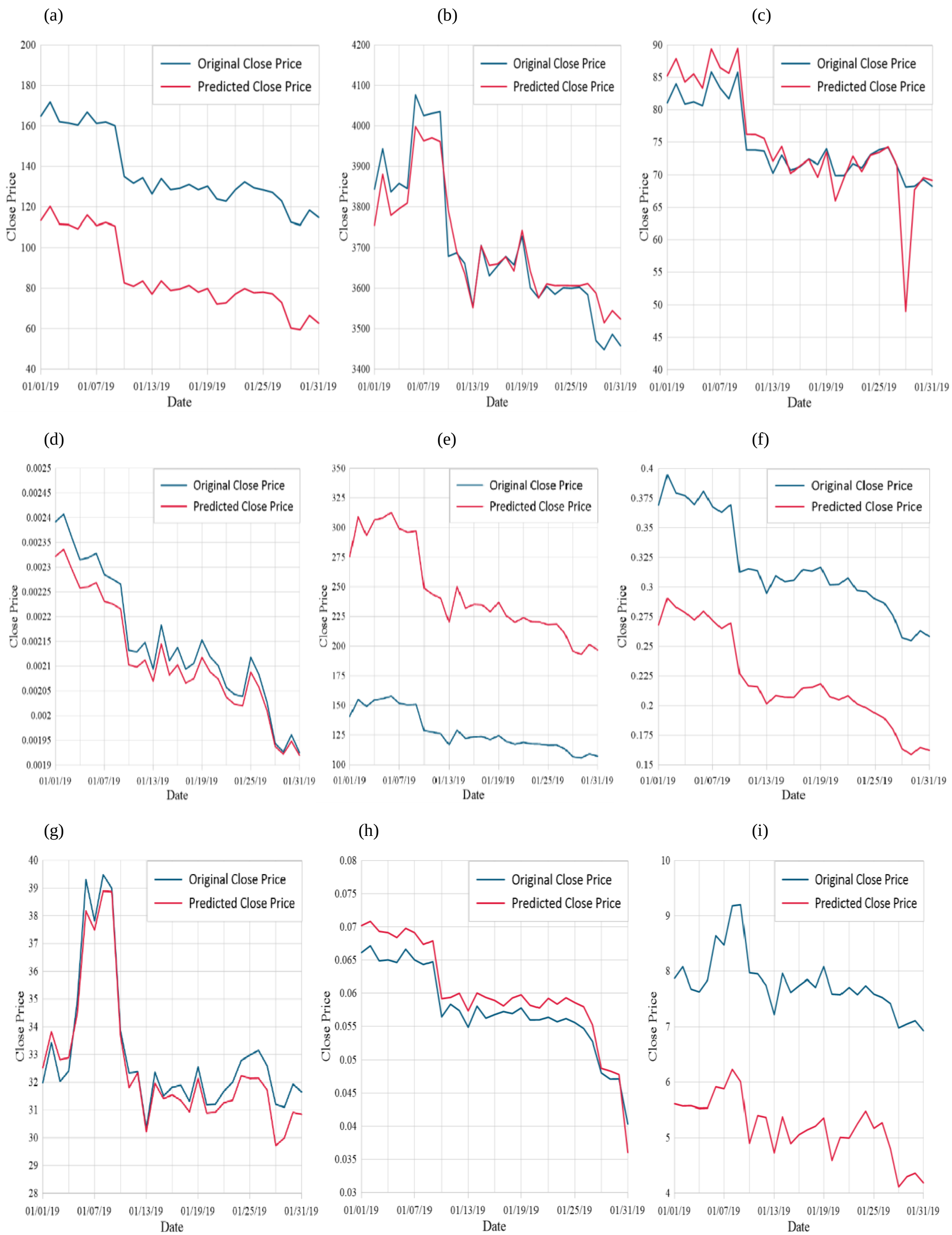


Figure 4. Comparison between original and predicted close price obtained from RapidMiner using neural net model for the month January 2019 (a) of constituent Bitcoin Cash. (b) of constituent Bitcoin. (c) of constituent Dash. (d) of constituent Dogecoin (DOGE). (e) of constituent Ethereum. (f) of constituent IOTA (MIOTA). (g) of constituent Litecoin. (h) of constituent NEM. (i) of constituent NEO.

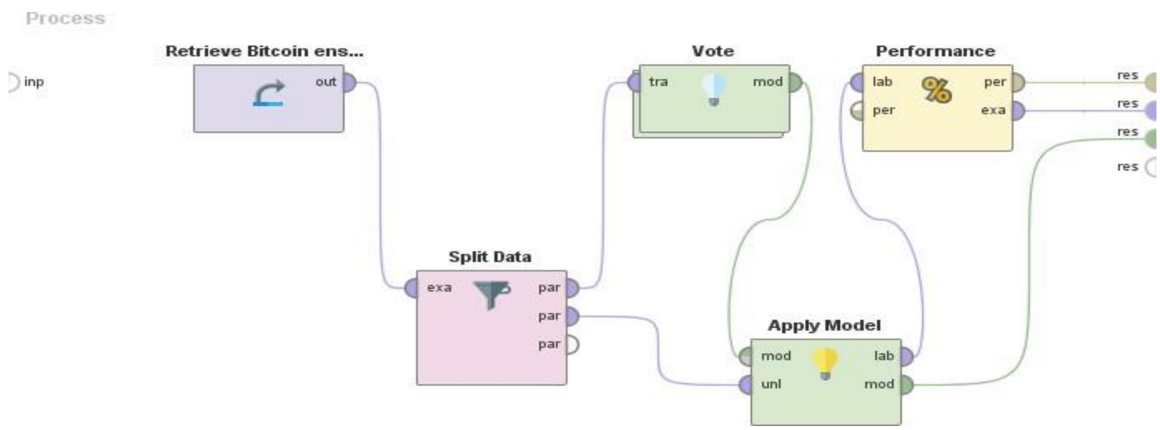
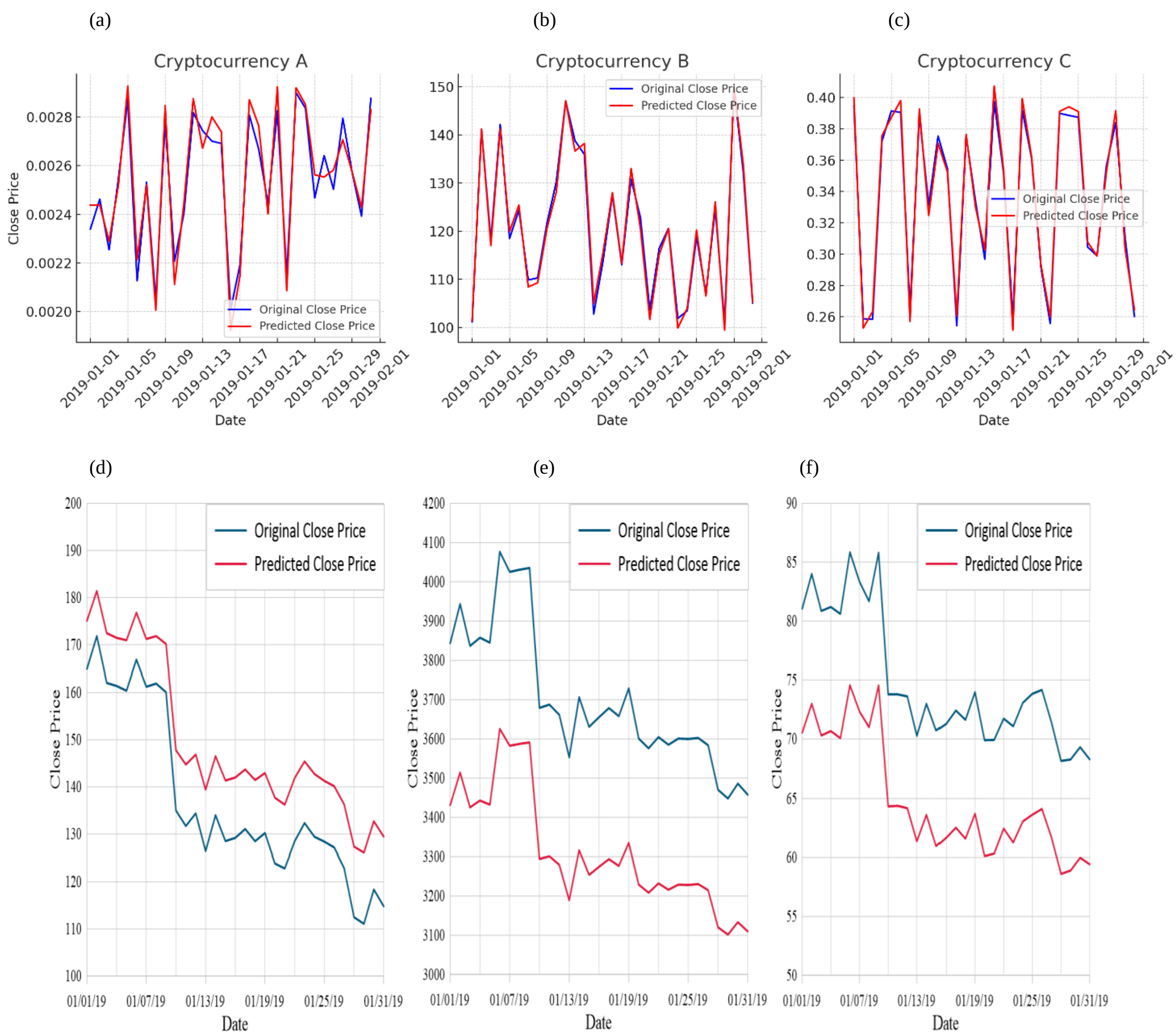


Figure 5. Ensemble learning model for predicting constituents and index of cryptocurrencies.



International Journal of Research Publication and Reviews, Vol (5), Issue (5), May (2024) Page – 12479-12495

(g) (h) (i)

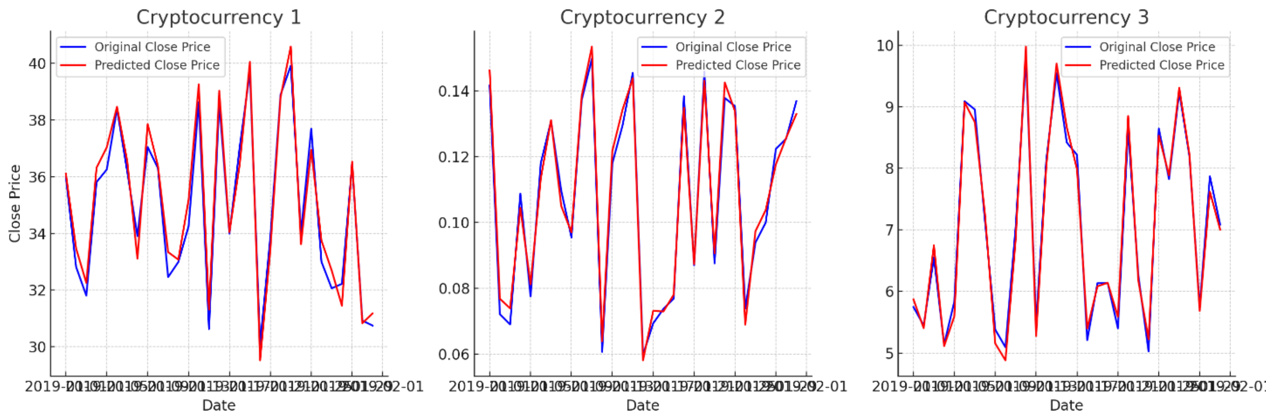


Figure 6. Comparison between original and predicted close price obtained from RapidMiner using ensemble learning method for the month January 2019 (a) of constituent Bitcoin Cash. (b) of constituent Bitcoin. (c) of constituent Dash. (d) of constituent Dogecoin (DOGE). (e) of constituent Ethereum. (f) of constituent IOTA (MIOTA). (g) of constituent Litecoin. (h) of constituent NEM. (i) of constituent NEO.

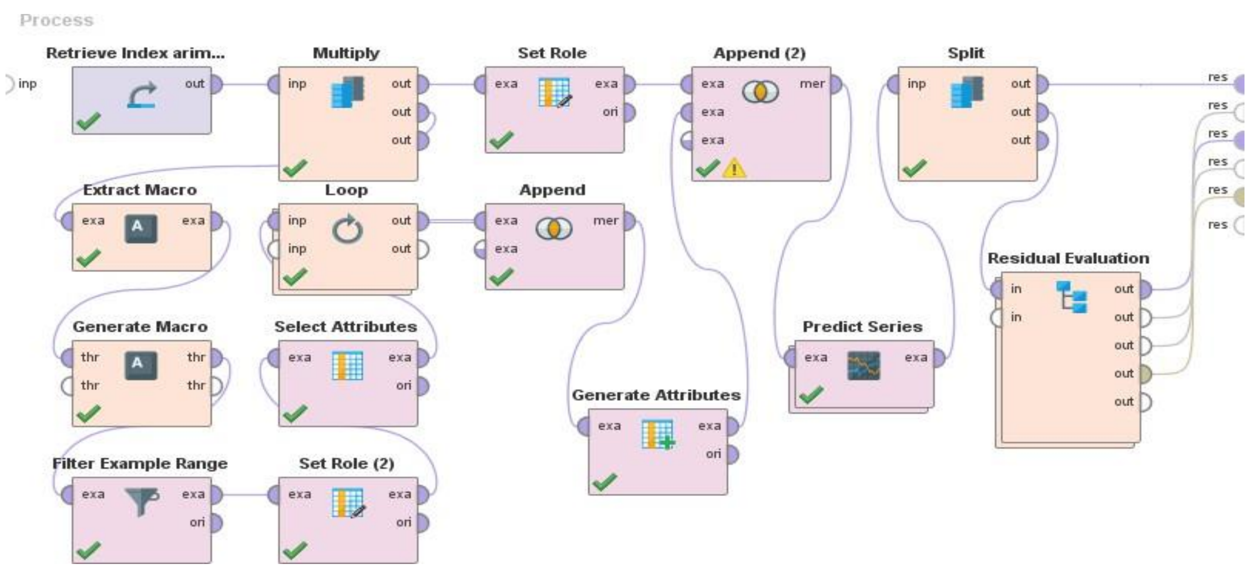
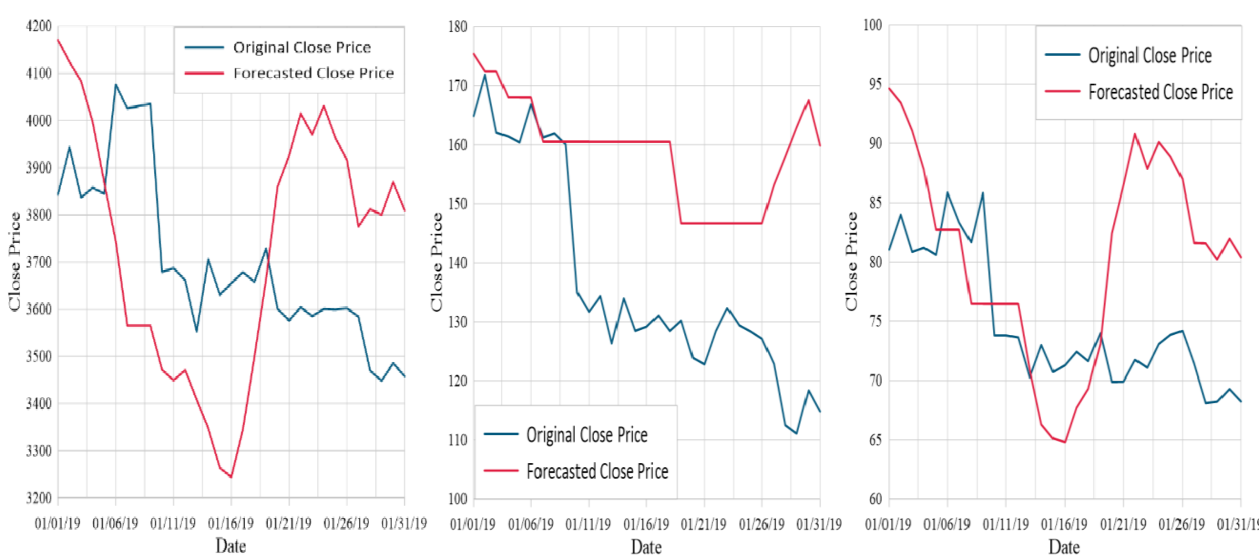


Figure 7. K-NN model for forecasting constituents and index of cryptocurrencies.

(a) (b) (c)



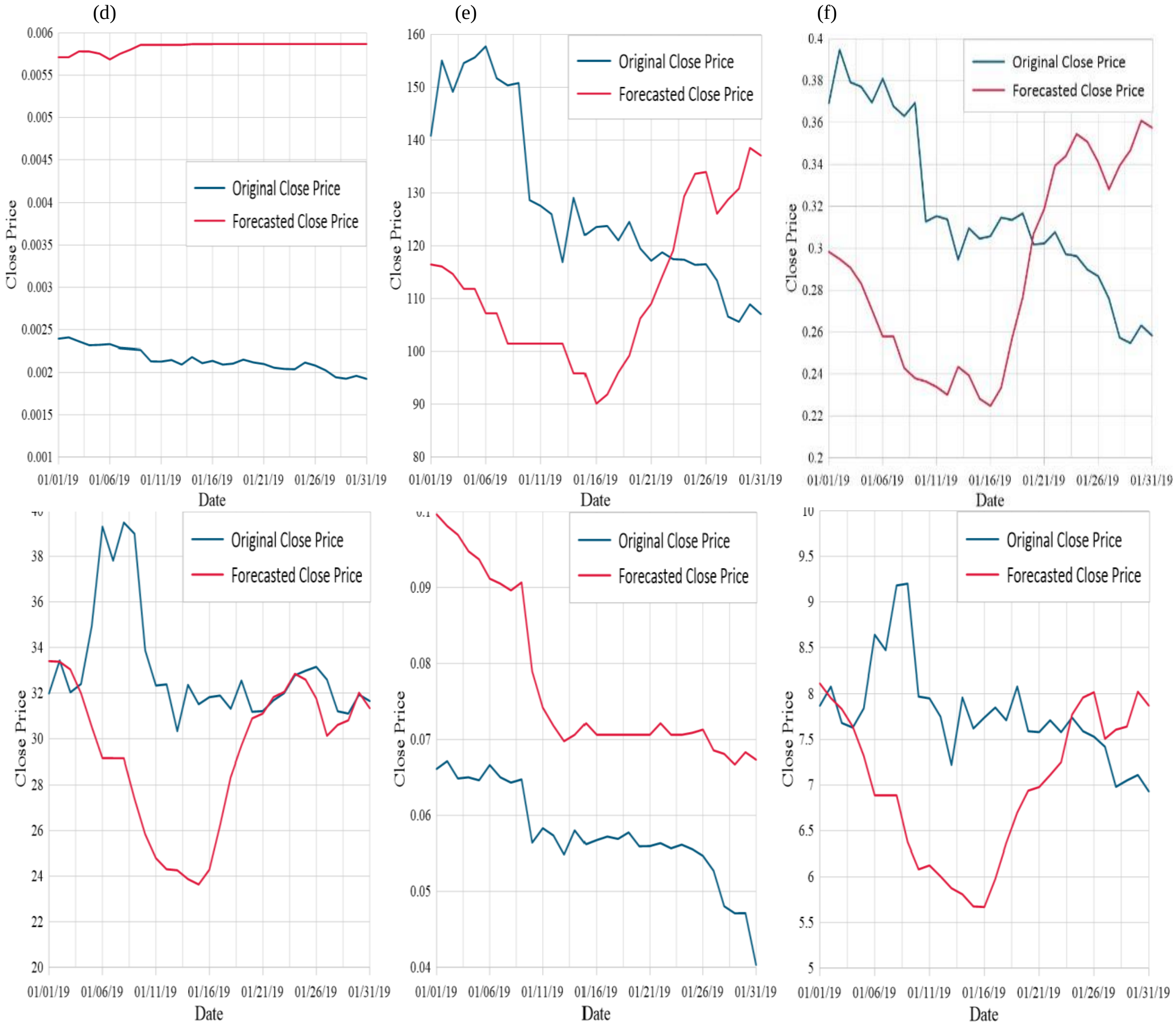
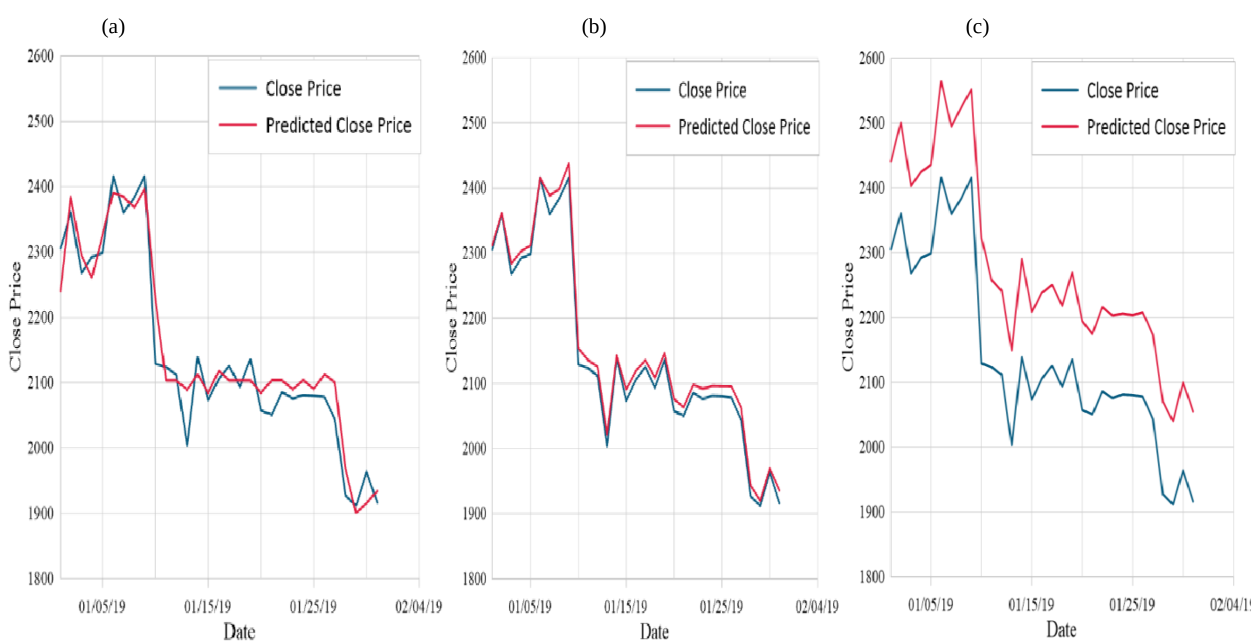


Figure 8. Comparison between original and forecasted close price obtained from RapidMiner using K-NN model (a) of constituent Bitcoin. (b) of constituent Bitcoin Cash. (c) of constituent Dash. (d) of constituent Dogecoin (DOGE). (e) of constituent Ethereum. (f) of constituent IOTA (MIOTA). (g) of constituent Litecoin. (h) of constituent NEM. (i) of constituent NEO.



(d)

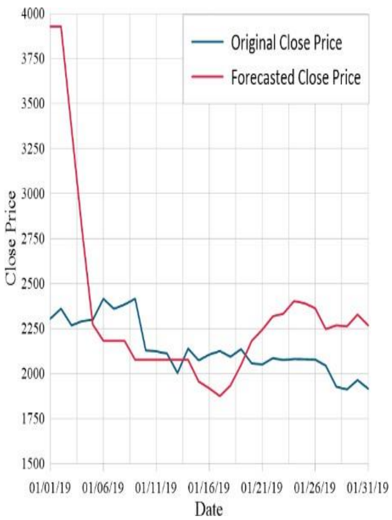
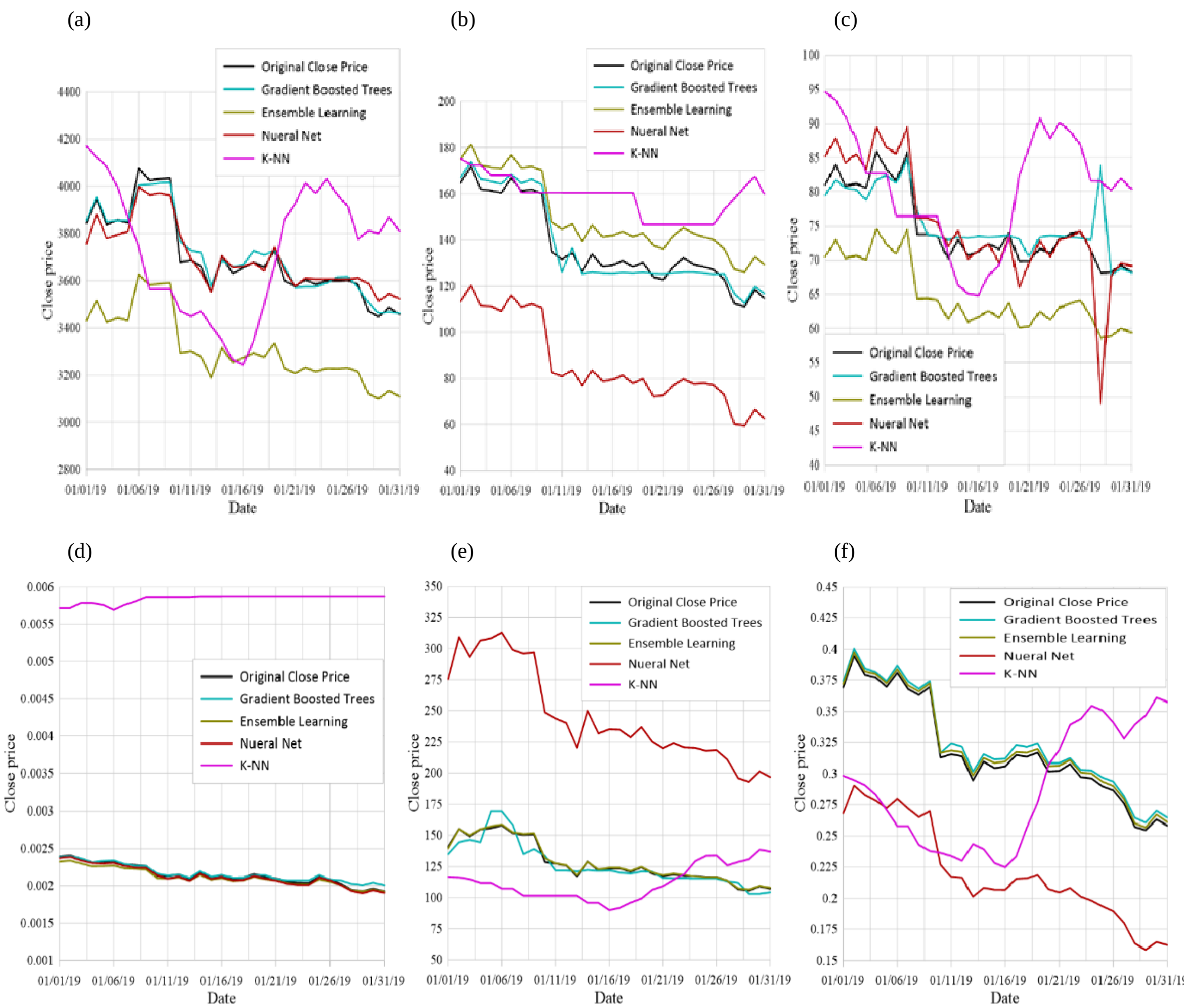
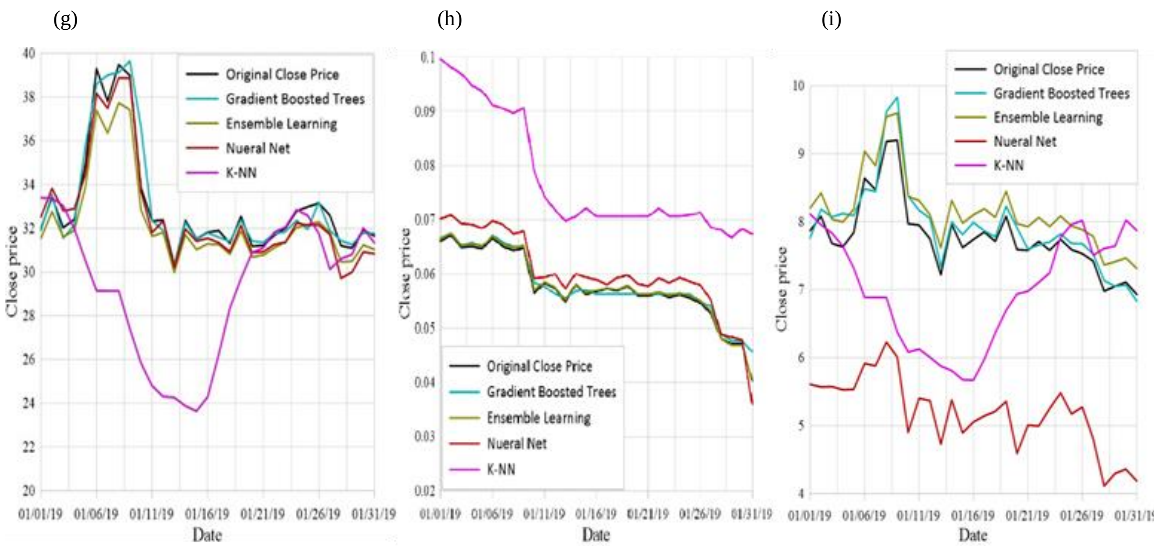


Figure 9. Comparison between original and forecasted close price obtained from RapidMiner of index cci30

(a) using gradient boosted trees model. (b) using ensemble learning method. (c) using neural net model (d) using K-NN model.





REFERENCE

[1] Choo, Kim-Kwang Raymond. “Cryptocurrency and Virtual Currency: Corruption and Money Laundering/ Terrorism Financing Risks?” Handbook of Digital Currency Bitcoin, Innovation, Financial Instruments, and Big Data, Elsevier, 2015, pp. 283–307.

[2] Deng, L., Che, J., Chen, H., & Zhang, L. (2018). Research on the Pricing Strategy of the CryptoCurrency Miner’s Market. Lecture Notes In Computer Science, 228-240.

[3] Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. Consulted (2008)

[4] Zhang, W., Wang, P., Li, X., & Shen, D. (2018). Some stylized facts of the cryptocurrency market. Applied Economics, 50(55), 5950-5965.

[5] GAINSBURY, Sally M., and ALEX BLASZCZYNSKI. Gaming Law Review. Vol. 21, Mary Ann Liebert, Inc., 2017.

[6] Barkatullah, J., & Hanke, T. (2015). Goldstrike 1: CoinTerra's First-Generation Cryptocurrency Mining Processor for Bitcoin. IEEE Micro, 35(2), 68-76.

[7] Li, X., & Wang, C. (2017). The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. Decision Support Systems, 95, 49-60.

[8] Cocco, L., Concas, G., & Marchesi, M. (2015). Using an artificial financial market for studying a cryptocurrency market. Journal Of Economic Interaction And Coordination, 12(2), 345-365.

[9] Ha, S., & Moon, B. (2018). Finding attractive technical patterns in cryptocurrency markets. Memetic Computing, 10(3), 301-306.

[10] Gkillas, K., Bekiros, S., & Siriopoulos, C. (2018). Extreme Correlation in Cryptocurrency Markets. SSRN Electronic Journal.

[11] Stosic, D., Stosic, D., Ludermir, T., & Stosic, T. (2018). Collective behavior of cryptocurrency price changes. Physica A: Statistical Mechanics And Its Applications, 507, 499-509.

[12] Alessandretti, L., ElBahrawy, A., Aiello, L., & Baronchelli, A. (2018). Anticipating Cryptocurrency Prices Using Machine Learning. *Complexity*, *2018*, 1-16.

[13] Sun Yin, H., & Vatrapu, R. (2017). A first estimation of the proportion of cybercriminal entities in the bitcoin ecosystem using supervised machine learning. 2017 IEEE International Conference on Big Data (Big Data).

[14] J. Han, J. C. Rodriguez and M. Beheshti, "Diabetes Data Analysis and Prediction Model Discovery Using RapidMiner," 2008 Second International Conference on Future Generation Communication and Networking, Hainan Island, 2008, pp. 96-99.

[15] Sun, X., Liu, M., & Sima, Z. (2018). A novel cryptocurrency price trend forecasting model based on LightGBM. *Finance Research Letters*.

[16] Lahmiri, S., & Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, *118*, 35-40.

[17] Kim, Y. B., Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., & Kim, C. H. (2016). Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. *PLOS ONE*, *11*(8), e0161197.

[18] Greaves, A., & Au, B. (2019). Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin.

[19] Barnwal, A., Bharti, H. P., Ali, A., & Singh, V. (n.d.). Stacking with neural network for cryptocurrency investment.

[20] McNally, S., Roche, J., & Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. *2018 26th Euromicro International Conference on Parallel, Distributed and Network- based Processing (PDP)*.

[21] Bakar, N. A., & Rosbi, S. (2017). Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency

Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction. *International Journal of Advanced Engineering Research and Science*, *4*(11), 130-137.

[22] Rebane, J., Karlsson, I., Denic, S., & Papapetrou, P. (2018). Seq2Seq RNNs and ARIMA models for Cryptocurrency Prediction: A Comparative Study. In SIGKDD Fintech'18. London: Association for Computing Machinery (ACM).

[23] Jiang, Zhengyao, and Jinjun Liang. "Cryptocurrency Portfolio Management with Deep Reinforcement Learning." 2017 Intelligent Systems Conference (IntelliSys), 2017.

[24] Kai-Sang Leung, C., Kyle Mackinnon, R., & Wang, Y. (2014). A machine learning approach for stock price prediction. In IDEAS '14 Proceedings of the 18th International Database Engineering & Applications Symposium (pp. 274-277). Porto: ACM New York, NY, USA ©2014.

[25] k, T., & Wadhawa, M. (2016). Analysis and Comparison Study of Data Mining Algorithms Using Rapid Miner. *International Journal Of Computer Science, Engineering And Applications*, *6*(1), 9- 21.

[26] Yin, Haohua Sun, and Ravi Vatrapu. "A First Estimation of the Proportion of Cybercriminal Entities in the Bitcoin Ecosystem Using Supervised Machine Learning." *2017 IEEE International Conference on Big Data (Big Data)*, 2017.

[27] Guo, T., Bifet, A., & Antulov-Fantulin, N. (2018). Bitcoin Volatility Forecasting with a Glimpse into Buy and Sell Orders. *2018 IEEE International Conference on Data Mining (ICDM)*.

[28] Jang, H., & Lee, J. (2018). An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. *IEEE Access*, *6*, 5427-5437.

[29] Indera, N., Yassin, I., Zabidi, A., & Rizman, Z. (2018). Non-linear Autoregressive with Exogeneous input (narx) bitcoin price prediction model using PSO-optimized parameters and moving average technical indicators. *Journal of Fundamental and Applied Sciences*, *9*(3S), 791.

[30] Fischer, T., Krauss, C., & Deinert, A. (2019). Statistical Arbitrage in Cryptocurrency Markets. *Journal Of Risk And Financial Management*, *12*(1), 31.

[31] Tahir, R., Huzaifa, M., Das, A., Ahmad, M., Gunter, C., Zaffar, F., … Borisov, N. (2017). Mining on Someone Else's Dime: Mitigating Covert Mining Operations in Clouds and Enterprises. *Research in Attacks, Intrusions, and Defenses*, 287-310.

[32] Kim, B. Y., Choi, S. S., & Jang, J. W. (2018). Data Managing and Service Exchanging on IoT Service Platform Based on Blockchain with Smart Contract and Spatial Data Processing. *Proceedings of the 2018 International Conference on Information Science and System - ICISS '18*.

[33] *Building a Sales Forecasting Model with RapidMiner* [Video file]. (2017, April 12). Retrieved from https://www.youtube.com/watch?v=_mSRr2aIZuw

[34] *Cryptocurrency Index 30 | CCi30*. (n.d.). Retrieved from https://cci30.com/